

PATENTTI- JA REKISTERIHALLITUS  
NATIONAL BOARD OF PATENT AND REGISTRATION

Helsinki 20.10.98

PCT/FI98/00749

E T U O I K E U S T O D I S T U S  
P R I O R I T Y D O C U M E N T

REC'D 23 NOV 1998  
WIPO PCT



Hakija FINNISH IMMUNOTECHNOLOGY LTD  
Applicant Tampere

Patentihakemus nro 973762  
Patent application no

Tekemispäivä 23.09.97  
Filing date

Kansainvälinen luokka C 12N  
International class

Keksinnön nimitys  
Title of invention

"Novel gene"  
(Uusi geeni)

**PRIORITY  
DOCUMENT**

SUBMITTED OR TRANSMITTED IN  
COMPLIANCE WITH RULE 17.1(a) OR (b)

Täten todistetaan, että oheiset asiakirjat ovat tarkkoja  
jäljennöksiä patentti- ja rekisterihallitukselle alkuaan  
annetuista selityksestä, patenttivaatimuksista, tiivistelmästä  
ja piirustuksista.

This is to certify that the annexed documents are true copies  
of the description, claims, abstract and drawings originally  
filed with the Finnish Patent Office.

*Satu Vasenius*  
Satu Vasenius  
Jaostopäällikkö

Maksu 385,- mk  
Fee 385,- FIM

Osoite: Arkadiankatu 6 A Puhelin: 09 6939 500 Telefax: 09 6939 5204  
Address: P.O.Box 1160 Telephone: + 358 9 6939 500 Telefax: + 358 9 6939 5204  
FIN-00101 Helsinki, FINLAND

1

2

**Novel gene****Field of the invention**

The present invention relates to a novel gene, a  
5 novel protein encoded by said gene, a mutated form of the  
gene and to diagnostic and therapeutic uses of the gene or  
a mutated form thereof. More specifically, the present  
invention relates to a novel gene defective in autoimmune  
polyendocrinopathy syndrome type I (APS I), also called  
10 autoimmune polyendocrinopathy-candidiasis-ectodermal  
dystrophy (APECED) (MIM No. 240,300).

**Background**

Autoimmune polyglandular syndrome type I (APS I),  
also known as autoimmune polyendocrinopathy-candidiasis-  
15 ectodermal dystrophy (APECED), is a rare recessively  
inherited disease (MIM No. 240,300) that is more prevalent  
among certain isolated populations, such as Finnish,  
Sardinian and Iranian Jewish populations. The incidence of  
the disease among the Finns and the Iranian Jews is esti-  
20 mated to be 1:25000 and 1:9000, respectively, whereas only  
few cases in other parts of the world are found each year.

APECED is one of the two major autoimmune poly-  
endocrinopathy syndromes. The causing factor of APECED has  
not yet been identified. In APECED, the patient develops  
25 chronic mucocutaneous candidiasis soon after birth, and  
later several organ-specific autoimmune diseases, mainly  
hypoparathyreoidism, Addison's disease, chronic atrophic  
gastritis with or without pernicious anemia, and in puberty  
gonadal dysfunction occur [Ahonen P, Clin. Genet. 27 (1985)  
30 535-542]. An accepted criterion for diagnosis of APECED is  
the presence of at least two of the three main symptoms,  
Addison's disease, hypoparathyroidism and candidiasis, in  
patients [Neufeld, M. et al., Medicine 60 (1981) 355-362].  
Immunologically, the major findings are the presence  
35 of high-titer serum autoantibodies against the ef-  
fected organs, antibodies against *Candida albicans*, and

low or lacking T-cell responses toward candidal antigens [Blizzard, R. M. and Kyle M., J. Clin. Invest. 42 (1963) 1653-1660; Arulanantham, K. et al., New Eng. J. Med. 300 (1979) 164-168; Krohn, K. et al., Lancet 339 (1992) 770-773; Uibo R. et al., J. Clin. Endocrinol. Metab. 78 (1994) 323-328]. The disease usually occurs in childhood, but new tissue specific symptoms may appear throughout life [Ahonen, P. et al., New Engl. J. Med. 322 (1990) 1829-1836]. APECED is not associated with a particular HLA haplotype, and both males and females are equally affected consistent with the autosomal recessive mode of inheritance.

The locus for the APECED gene has been mapped to chromosome 21q22.3 between gene markers D21S49 and D21S171 based on linkage analysis of Finnish families [Aaltonen, J. et al., Nature Genet. 8 (1994) 83-87]. Recently, Börses et al. reported a maximum LOD score of 10.23 with marker D21S1912 just proximal to the gene PFKL, and thus by linkage disequilibrium studies the critical region for APECED can be considered to be less than 500 kb between markers D21S1912 and D21S171. Locus heterogeneity was not revealed by linkage analysis of non-Finnish families [Björsses, P. et al., Am. J. Hum. Genet. 59 (1996) 879-886].

Physical maps of human chromosome 21q22.3 have been developed using YACs, and bacterial based large insert cloning vectors [Chumakov et al., Nature 359 (1992) 380; Stone et al., Genome Res. 6 (1996) 218], and many laboratories have contributed to the construction of a transcription map of the whole chromosome and 21q22.3 in particular [Chen et al., Genome Res. 6 (1996) 747-760; Yaspo et al., Hum. Mol. Genet. 4 (1995) 1291-1304]. Numerous trapped exons from chromosome 21 specific cosmids and also physical contigs from the APECED critical region have been identified and partially characterized. In addition, a number of ESTs from the international human

genome project have been mapped to the APECED critical region.

Recently, as part of the international efforts of generating the entire sequence of human chromosome 21 and 5 international agreements on the immediate availability of this type of sequence data, the partial sequence of the APECED gene critical region was made available in GenBank by the Stanford Human Genome Center which is currently carrying out the sequencing of 1.0 Mb around the critical 10 region of the APECED gene.

However, the precise location and the sequence of the APECED gene and the nature of the gene product have not so far been clarified. Thus at present the diagnosis of APECED is based mainly on developed clinical symptoms and 15 typical clinical findings, e.g. the presence of autoantibodies against adrenal cortex or steroidogenic enzymes P450c17 and/or P450scc. The linkage analysis is seldom used. Further, means for natal or presymptomatic diagnosis of the disease are not easily available, since the linkage 20 analysis provides only an indirect data through known gene markers and requires samples from several family members in several generations. Additionally, the linkage analysis is tedious and can be performed only in specialized laboratories by highly-skilled personnel.

25 Also the mapping of the carriers of the disease gene is presently based on the linkage analysis and thus not readily available.

#### **Summary of the invention**

We have now identified a novel gene encoding a novel 30 zinc finger protein, designated as autoimmune regulator 1 or AIR-1, which is mutated in APECED. The novel gene and protein allow further development of the diagnosis and therapy of APECED.

The object of the invention is to provide means 35 which are useful in a diagnostic method and a gene therapeutic method in the diagnosis and treatment of APECED.

Another object of the invention is to provide a novel method for the diagnosis APECED, including the pre- and postnatal diagnosis of and the mapping of the carriers, the method being easy and reliable to perform.

5 The present invention relates to an isolated DNA sequence comprising the sequence id. no. 1 or a fragment or variant thereof, or an isolated DNA sequence hybridizable thereto, the DNA sequence being associated with APECED. Preferably said isolated DNA sequence includes a gene 10 defect responsible for APECED.

The present invention also relates to a protein comprising the amino acid sequence id. no. 2 or a fragment or variant thereof, the protein being associated with APECED. Said protein has distinct structural motifs, 15 including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich region (PRR), and cysteine-rich region (CRR).

The present invention further relates to a method for the diagnosis of APECED comprising detecting in a biological specimen the presence of a DNA sequence 20 comprising the sequence id. no. 1 or a functional fragment or variant thereof, or a DNA-sequence hybridizable thereto, the DNA sequence being associated with APECED.

The present invention further relates to the use of the above-identified DNA-sequences in the diagnosis of 25 APECED.

The present invention further relates to a method for the diagnosis of APECED comprising detecting in a biological specimen the presence or the absence of a protein comprising the sequence id. no. 2 or a fragment 30 thereof, the protein being associated with APECED.

The present invention further relates to the use of the above-identified protein or a fragment thereof in the diagnosis of APECED.

The present invention further relates to the use of 35 the above-identified DNA sequences in gene therapy or for

the preparation of a pharmaceutical preparation useful in a gene therapy method of APECED.

**Brief description of the drawings**

Figure 1 shows a physical map of the APECED gene locus in the chromosome 21q22.3. Cosmids D1G8, D40G11, 5 D9G11, D28B11, and D4G11, overlapping clones used for the genomic sequencing [Kudoh, J. et al., DNA Res. 4 45 -52] are indicated by horizontal lines. The APECED gene located just proximal to the 5' end of the neighboring gene 10 PFKL is indicated by a solid arrow. N indicates *NotI* sites. DNA marker D21S1912 is shown as open box.

Figure 2 shows the structures of the APECED gene and AIR proteins. (A) Cloning strategy of AIR cDNAs and the order of the exons in the APECED gene. DNA fragments 15 amplified by PCR and 3'- and 5'-RACE are indicated by the lines. Exon 1' is the 5'-noncoding exon of the AIR-2 and AIR-3. An additional alternative splicing of AIR-3 in exon 10, resulting in an amino acid change in its downstream, is indicated by vertical lines. Each exon, except exon 1', is 20 bordered by the common splice site consensus sequence, ag:gt. Mutations in the exon 2 and exon 6 are indicated by the arrows. (B) Schematic presentation of the three AIR proteins showing distinct structural motifs, including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich 25 region (PRR), and cystein-rich region (CRR).

Figure 3 shows electropherograms showing the sequence surrounding the mutations in the APECED gene. (A) Mutation analysis of a Swiss APECED family. The parents are heterozygous for the allele (normal "C" and abnormal "T"). 30 The affected boy and girl show the "C" to "T" transition resulting in the "Arg" to "Stop" nonsense mutation at amino acid position 257. (B) Mutation analysis of two Finnish APECED patients. The patient MP is homozygous for the mutant allele (left), NP is heterozygous for the allele 35 (right). (C) The patient NP shows the "A" to "G" trans-

version resulting in the "Lys" to "Glu" missense mutation at amino acid position 42. FLEB is a normal control.

Figure 4 shows the result of restriction enzyme *TaqI* digestion assay demonstrating the R257stop mutation. Four APECED patients [HP1 (lane 1), HP2 (lane 2), NP (lane 6), and MP (lane 8)], the mothers of two families [HM (lane 5) and NM (lane 7)], two healthy siblings [HN1 (lane 3) and HN2 (lane 4)] of family H and normal controls [C1, C2 and C3 (lanes 9-11)] are shown. The APECED patients HP1, HP2 and MP are homozygotes for R257stop mutation. The APECED patient NP is heterozygous for R257stop mutation but is carrying a mutation at a different position in another allele of APECED gene (shown above in Fig. 3C). Both mothers (HM and NM) and two healthy siblings (HN1 and HN2) are heterozygous for R257stop mutation and therefore carriers of APECED but are not having the disease. Two controls (C1 and C2) are both homozygous for normal alleles. Normal alleles produce a lower 225 bp fragment, the mutated fragment is upper band at 285 bp.

Figure 5 shows an amino acid sequence alignment for the PHD finger motif of AIR-1, Mi-2, and TIF1. The consensus amino acid residues conserved in the PHD finger motif is indicated by the bold letters underneath. The residues that are identical with AIR-1 (aa 299-340) are shown by the dots. GenBank accession nos. of Mi-2 and TIF1 are X86691 and AF009353, respectively.

Figure 6. A Western blot showing the expression of AIR-1 in fetal liver. A sample of fetal liver was run on PAGE, transferred to nitrocellulose filter and probed with sera as follows: Lane 1 control mouse serum, lane 2, control mouse serum absorbed with peptide AIR-1/2 (sequence id. no. 25), lanes 3 and 4, serum from a mouse immunized with peptide AIR-1/2 for four and six weeks, respectively and absorbed with peptide AIR-1/2, lanes 5 and 6 unabsorbed serum from a mouse immunized with peptide AIR-1/2 for four

and six weeks, respectively. The strong band seen in lanes 5 and 6 represent the AIR-1 protein with a molecular weight of approx. 58 kD, the lower band is an approx. 20 kD breakdown product of the AIR protein. The bands seen in all 5 lanes are non-specific.

#### Detailed description of the invention

The present invention is based on studies aiming for the identification and characterization of the gene defect in APECED. In the sequence studies, a cosmid/BAC (bacterial artificial chromosome) contig of 520 kb covering four gene markers D21S1460-D21S1912-PFKL-D21S154 [Kudoh, J. et al., DNA Res. 4 (1997) 45-52] was constructed, and genomic sequencing in this region was performed [Kawasaki, K. et al., Genome Res. 7 (1997) 250-261]. From this genomic sequence information the distance between D21S1912 and PFKL was determined to be approximately 140 kb (Fig. 1).

Using a computer program, such as GRAIL and GENSCAN [Uberbacher, E. C. and Mural, R. J., Proc. Natl Acad. Sci. USA 88 (1991) 11261-11265; Burge, C. and Karlin, S., J. Mol. Biol. 268 (1997) 78-94], gene screening in the partial sequencing data within this region was performed. GENSCAN predicted several genes between D21S1912 and PFKL. One of these genes located just proximal to the PFKL gene contained the previously trapped exon HC21EXc33 [Kudoh, J. et al., DNA Res. 4 (1997) 45-52] or MDC04M06 [Chen, H. et al., Genome Res. 6 (1996) 747-760]. A set of primers for polymerase chain reaction (PCR) was then designed from the predicted exons. The PCR screening of various cDNA libraries using these primers allowed the isolation of a cDNA clone containing the exon HC21EXc33 (exon 13) from the thymus cDNA library (Fig. 2A).

A 3'-rapid amplification of cDNA ends (3'-RACE) and 5'-RACE using Marathon™ cDNA Amplification Kit (Clontech Laboratories Inc, California, USA) according to manufacturer's protocol from the thymus cDNA library was

performed using a primer c33F (sequence id. no. 7) and a primer 1R (sequence id. no. 8), respectively.

Sequencing analysis revealed a unique sequence of 2027 bp in overlapping PCR products that contains a 1635-bp open reading frame (ORF) from methionine at nt 128 to a TAG stop codon at nt 1763 encoding a predicted novel protein designated AIR-1, for autoimmune regulator 1. AIR-1 encodes a protein of 545 amino acids with a predicted isoelectric point of 7.32 and a calculated molecular mass of 57,723 (Fig. 2B).

A 5'-RACE from the thymus cDNA using a primer 4R (sequence id. no. 9) resulted in an alternatively spliced product. Furthermore, two types of the cDNA clones were amplified with a primer pair 3F/c33R (sequence id. no. 10/sequence id. no. 11) and these clones encode for AIR-2 and AIR-3 proteins sequence id. no. 4 and sequence id. no. 6, respectively (Fig. 2A) (sequence id. no. 3 and sequence id. no. 5). The AIR-2 and AIR-3 proteins consist of 348 and 254 amino acids, respectively (Fig. 2B). These results suggest that the APECED gene is transcribed as at least three types of mRNA by alternative splicing and/or use of an alternative 5' exon within the gene. RT-PCR analysis [Griffin, H. G. and Griffin, A. M., PCR Technology. Current Innovations, CRC Press, 1994] revealed that the AIR-1 transcript is also expressed in fetal liver (data not shown).

The APECED gene is approximately 13-kb in length and contains 15 exons, including the exon 1' specific to AIR-2 and AIR-3. It is transcribed in the direction of centromere to telomere (Figs 1, 2A). Based on this information, PCR primers were designed to amplify each exon from the genomic DNA and a mutation analysis of Swiss and Finnish APECED families was performed. Sequence comparison identified two mutations in the APECED gene of the patients (Fig. 3). The first mutation changes an Arg codon (CGA) to a stop codon (TGA) at amino acid position 257 in exon 6.

This mutation was designated as R257stop mutation. The second mutation is a missense mutation that derived from the maternal chromosome in one Finnish patient (NP): a Lys codon (AAG) changes to a Glu codon (GAG) at amino acid 5 position 42 in exon 2. This mutation is designated as L42E mutation (Figs 2A, 3C).

The R257stop mutation destroys a *TaqI* restriction enzyme site and the K42E mutation introduces a novel *TaqI* site. Thus these two mutations can be easily demonstrated 10 in one or both alleles by *TaqI* digestion or by digestion using another enzyme cleaving at the recognition site 5'-TCGA-3' (Fig. 4).

The AIR-1 protein has strong homology in certain domains to the major autoantigens (Mi-2) associated with the 15 autoimmune disease dermatomyositis [Seeig, H. P. et al., *Arthritis Rheum.* 38 (1995) 1389-1399; Ge, Q. et al., *J. Clin. Invest.* 96 (1995) 1730-1737], Sp140, a protein from the nuclear body, an organelle involved in the pathogenesis 20 of certain types of leukemia, and which is also the target of antibodies in the serum of patients with the autoimmune disease primary biliary cirrhosis [Bloch, D. B. et al., *J. Biol. Chem.* 271 (1996) 29198-29204]. In addition, the homologies extend to other nuclear proteins such as TIF1 25 [Le Douarin, B. et al., *EMBO J.* 14 (1995) 2020-2033], LYSP100 [Dent, A. L. et al., *Blood* 88 (1996) 1423-1426], and putative yeast and *C. elegans* proteins. The AIR-1 protein homologies are principally in two PHD finger motifs (amino acid 299 to 340 and 434 to 475) (Fig. 5). AIR-1 also contains a proline-rich regions (amino acid 350 to 430) 30 (Fig. 2B). The PHD finger is a cysteine-rich structure that is distinguished from the RING finger (C3HC4) and LIM domain (C2HC5) because it contains a consensus of C4HC3. [Aasland, R. et. al., *Trends Biochem. Sci.* 20 (1995) 56-59]. The PHD finger motif is found in a number of 35 chromatin-associated proteins such as HRX that is involved in the t(11:17) translocation in acute leukemia [Chaplin,

T. et al., Blood 86 (1995) 2073-2076]. The proline-rich region is assumed to be involved in protein-protein interaction or DNA binding. The presence of the PHD finger and proline-rich regions indicates a function for AIRs as transcription regulatory proteins. However, the AIR proteins have no apparent nuclear translocation signal, and thus other proteins containing such signal may interact with AIR to translocate it to the nucleus. In fact, the AIR proteins also have the LXXLL motif that is a signature sequence to bind to nuclear receptors [Heery, D. M. et al., Nature 387 (1997) 733-736] (Fig. 2B).

The clinical picture of APECED and the observed immunological abnormality with strong autoimmune response towards several target organs and antigens suggest that the product of the APECED gene has a central role in immune (ontogeny) maturation and in regulation of immune response towards self and nonself.

According to the diagnostic method of the invention, the presence of the defective APECED gene can be detected from a biological sample by any known detection method suitable for detecting mutations. Such methods include the method described by Saiki et al. [Proc. Natl. Acad. Sci USA 86 (1989) 6230-6234] utilizing hybridization to an allele specific oligonucleotide probe, or modifications thereof; the method described by Newton, C. R. et al. [Nucl. Acids Res. 17 (1989) 2503-2516] using the DNA sequences or DNA-fragments of the invention as probes; the solid phase minisequencing method described by Syvänen et al. [Genomics 8 (1990) 684-692] in which use is made of a biotinylated probe; or the oligonucleotide ligation method described by Landegren, U. et al. [Science 241 (1988) 1077-1080]. Methods include the denaturing gradient gel electrophoresis (DGGE) [Fischer, S.G. and Lerman, L.S., PNAS 80 (1983) 1579-1583] or a modification of this method, constant denaturant gel electrophoresis (CDGE) [Hoving et al., Genes Chromosomes Cancer 5 (1992) 97-103]. The mutation

separation principle of DGGE and CDGE is based on the melting behavior of the DNA double helix of a given fragment.

Since the mutations of the APECED gene involve 5 a site sensitive to *TaqI* digestion, the mutation are preferably detected in one or both alleles by *TaqI* digestion or by digestion using another enzyme cleaving at recognition site 5'-TCGA-3'. The chemical mismatch cleavage for mutation analysis can be used [Grompe, M. et al., Proc. 10 Natl. Acad. Sci. USA 86(15) (1989) 5888-5892].

In the diagnostic method of the invention the biological sample can be any tissue or body fluid containing cells, such as blood, e.g. umbilical cord blood, separated 15 blood cells, such as lymphocytes, B-cells, T-cells etc., biopsy material, such as fetal liver or thymus biopsy, sperm, saliva, etc. The biological sample can be, where necessary, pretreated in a suitable manner known to those skilled in the art.

When the DNA sequence of the present invention is 20 used therapeutically any techniques presently available for gene therapy can be employed. Accordingly, in the technique known as *ex vivo* therapy patient cells (e.g. umbilical cord blood from the fetus) with the defective gene are taken 25 from the patient, DNA sequences encoding the normal (healthy) gene product incorporated in a carrier vector are transduced or transfected to the cells and the cells are returned to the patient. If the techniques known as *in situ* therapy is used, the DNA sequences encoding the normal gene product are first inserted to a suitable carrier vector, 30 and the carrier is then introduced to the affected tissue, such as peripheral blood, liver or bone marrow. The carrier vector used can be a retrovirus vector, an adeno virus vector, an adeno associated virus (AAV) vector or an eucaryotic vector. The therapy can be performed intra utero 35 or during adult life. Depending on the cells to be treated these techniques lead either to a transient cure, where

cells from affected organ are treated, or to a permanent cure, in case of the treatment of stem cells.

The present invention provides means for an easy and more rapid diagnosis of the APECED and, specifically, 5 enables prenatal diagnosis and carrier diagnosis. Furthermore, it provides a background for therapy.

The invention is now elucidated by the following non-limiting examples.

Example 1

10                   **Localization of the APECED gene**

Genomic sequencing of cosmid DNAs was performed by the shotgun method described by Kawasaki, K. et al., Genome Res. 7 (1997) 250-261. Cosmids D1G8, D40G11, D9G11, D28B11, and D4G11 and gene marker D21S1912 are described by Kudoh, 15 J. et al., DNA Res. 4 (1997) 45-52].

CDNA cloning

The phage DNAs prepared from human thymus cDNA library (Clontech, HL1127a) were used as a PCR template. 20 ng of phage DNA which represents approximately  $4 \times 10^8$  phages was added to a 10 ml of reaction mixture containing 1x buffer [16mM  $(\text{NH}_4)_2\text{SO}_4$ , 50mM Tris-HCl, pH 9.2, 1.75 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin], 0.2 mM each of dNTPs, 1M Betaine (Sigma), 0.35 U of Tap and Pwo DNA polymerase (EXpand Long Template PCR System, Boehringer Mannheim), and 25 0.5 mM of each of the primers, 2F and c33R, 2F and 4R, and 2F' and 2R', respectively.

The cDNA fragment was amplified by PCR using the following conditions: 94°C for 3 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec in 2F/c33R and 2F/4R or 65°C for 30 sec in 2F'/2R', and 68°C for 90 sec. 3'- and 5'-RACE were carried out by Marathon cDNA Amplification Kit (Human Thymus; Clontech). PCR reaction was performed in 10 µl volume containing 1x buffer (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin), 0.2 mM each of 35 dNTPs, 0.25 U of AmpliTaq Gold polymerase (Perkin-Elmer), and 0.5 mM of each of the exon-specific primers. 3'-RACE

product was amplified by PCR with the following conditions: 95°C for 9 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 72°C for 30 sec.

The cDNA fragments were sequenced by the dye deoxy 5 terminator cycle sequencing method (according to ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit protocol P/N 402078, Perkin Elmer Corporation, California) using specific primers, 2F and c33R, and AmpliTaq/FS DNA polymerase (Perkin-Elmer), and then analyzed by using an 10 automatic DNA sequencer (Applied Biosystems 377). Primer sequences used were

1R: 5'-GTCCCCGAGTGGAAAGGCGCTGC-3' (sequence id. no. 8)  
2F: 5'-GGATTCAAGACCATGTCAGCTTCA-3' (sequence id. no. 12)  
3F: 5'-GAGTCAGGTACCCAGAGATGCTG-3' (sequence id. no. 10)  
15 c33R: 5'-CTCGCTCAGAAGGGACTCCA-3' (sequence id. no. 11)  
4R: 5'-AGGGGACAGGCAGGCCAGGT-3' (sequence id. no. 9)  
2F': 5'-GTGCTGTTCAAGGACTACAAC-3' (sequence id. no. 13)  
2R': 5'-TGGATGAGGATCCCCCTCCACG-3' (sequence id. no. 14)  
AP1: 5'-CCATCCTAATACGACTCACTATAAGGGC-3' (sequence id. no. 20 15) and  
c33F: 5'-GATGACACTGCCAGTCACGA-3' (sequence id. no. 7).

#### Example 2

##### **Mutation analysis of the APECED gene**

For the mutation analysis the DNA samples were 25 purified from periferal blood mononuclear cells from patients with APECED and from suspected carriers of APECED and from normal healthy controls (according to Sambrook et al. 1989, Molecular Cloning. A Laboratory Manual. CSH Press) and subjected to PCR using primers specific for all 30 identified exons.

For sequencing the mutated exons, PCR fragments, 6F/6R in exon 6 and 49300F/49622R in exon 2, were amplified by PCR with the following conditions: 95°C for 9 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec and 72°C for 30

sec, and 94°C for 3 min., 35 cycles of 94°C for 30 sec, 60°C for 30 sec, and 68°C for 30 sec, respectively. The PCR products were sequenced using specific primers

6F: 5'-TGCAGGCTGTGGAACTCCA-3' (sequence id. no. 16)

5 6R: 5'-AGAAAAAGAGCTGTACCTGTG-3' (sequence id. no. 17)

3R: 5'-TGCAAGGAAGAGGGCGTCAGC-3' (sequence id. no. 18)

49300F: 5'-TCCACCAACAAGCCGAGGAGAT-3' (sequence id. no. 19)

and 49622R: 5'-ACGGGCTCCTCAAACACCCT-3' (sequence id. no. 20).

10 In the mutation analysis by sequencing, two Swiss and three Finnish (HP1, HP2 and MP) patients with APECED were homozygous for R257stop allele, whereas one Finnish patient (NP) was heterozygous for this mutation (Fig. 3A, B). The R257stop mutation of NP was derived from the 15 paternal chromosome. The second mutation, L42E mutation, was found in one Finnish patient (NP): a Lys codon (AAG) changes to a Glu codon (GAG) at amino acid position 42 in exon 2. (Figs 2A, 3C). This mutation derived from the maternal chromosome.

20 Example 3

**Restriction enzyme TaqI analysis of two mutations in exons 2 and 6 of APECED gene**

Analysis of the mutation sites in exons 2 and 6 in a large series of individuals was performed using the 25 restriction enzyme TaqI. The TaqI digestion for exons 2 and 6 was done as follows. Ten microlitres of amplification product was incubated at 65 °C for 1 hour in 20µl of reaction mixture containing 1x TaqI digestion buffer (New England Biolabs, NY, 100 µl/ml of BSA and 10u of TaqI 30 enzyme (New England Biolabs, NY)). After the digestion fragments were separated in 1,5% agarose gel and visualized by EtBr staining.

For exon 2, the fragment containing the mutation site L42E was amplified with primers GR1/2F and GR1/2R with 35 the following conditions: 95°C for 3 min., 35 cycles of 94°C

for 30 sec, 62°C for 30 sec and 72°C for 1 min. The 1x reaction mix used contained 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM MgCl<sub>2</sub>, 0.001% (w/v) gelatin), 0.2 mM each of dNTPs, 0.25 U of Dynazyme (Finnzymes, Finland), and 0.5 mM 5 of each of the exon-specific primers. The normal allele produces a 312 bp fragment whereas the mutated allele gives a 133 bp and a 179 bp fragment. Primer sequences for GR1/2F and GR1/2R are 5'-TGGAGATGGGCAGGCCGCAGGGTG (sequence id. no. 21) and 5'-CAGTCCAGCTGGGCTGAGCAGGTG (sequence id. 10 no. 22), respectively.

For exon 6, the fragment containing the R257stop mutation site was amplified with primers GR1/5IF and GR1/5IR with the same conditions described for exon 2 (see above). The normal allele produces a 225 bp fragment 15 whereas the mutated allele gives a 285 bp fragment. Primer sequences for GR1/5IF and GR1/5IR are 5'GCGGCTCCAAGAAGTG-CATCCAGG (sequence id. no. 23) and 5'-CTCCACCCCTGCAAGGAA-GAGGGGC (sequence id. no. 24), respectively.

The screening of 50 Finnish and 50 Swiss healthy 20 individuals did not reveal R257stop or K42E mutations by *TaqI* digestion. Similarly, PCR analysis of 20 unaffected Japanese was performed and no mutations were found in any of these positions. These results demonstrate that the APECED gene is responsible for the pathogenesis of APECED.

Mutations were found in the AIR-1 transcript but not 25 in the AIR-2 and AIR-3 transcripts from all the APECED patients tested. Two Swiss and three Finnish (HP1, HP2 and MP) patients who are homozygous for the R257stop mutation completely lack functional AIR-1 protein but still have 30 intact AIR-2 and AIR-3 proteins.

One common mutation seems responsible for the genetic defect in approximately 90% of the Finnish APECED cases and a haplotype analysis with the markers D21S141, D21S1912 and PFKL shows that the R257stop mutation is

likely to be this common mutation [Björses, P. et al., Am. J. Hum. Genet. 59 (1996) 879-886].

**Example 4**

**Analysis of the AIR protein expression**

5 In this example, synthetic peptides representing amino-acid sequences of the AIR-1 protein, were used to generate a polyvalent mouse antiserum against the AIR-1 protein.

For the peptide synthesis, two peptides were chosen  
10 according to the antigenicity prediction by Pepsort program (GCC package, Wisconsin, USA). The peptides AIR-1/2 and AIR-1/6 (TLHLKEKEGCPQAFH, sequence id. no. 25 and GKNKARSSSGPKPLV, sequence id. no. 26, respectively) representing exons 2 and 6, respectively, of the APECED gene  
15 were synthesized onto a branched lysine core (Fmoc8-Lys4-Lys2-Lys-betaAla-Wang resin, Calbiochem-Novabiochem, La Jolla, Ca, USA) resulting in an octameric multible antigen peptide (MAP) [Tam, J. P. et al., Proc. Natl. Acad. Sci. USA 85 (1988) 5409-5413; Adermann, K. et al., in Solid  
20 Phase Synthesis, Biological and Biomedical Applications, pp. 429-432, Ed. R. Epton, Mayflower Worldwide Ltd., Birmingham, 1994], Syntheses were performed by Fmoc (N-(9-fluorenyl)methoxycarbonyl) chemistry on a simultaneous multiple peptide synthesizer (SMPS 350, Zinsser Analytic,  
25 Frankfurt, Germany). Purity of MAPs was analyzed by reverse-phase HPLC (System Gold, Beckman Instruments Inc, Fullerton, CA, USA).

To obtain murine polyclonal antibodies, eight-week old Balb/c mice were immunized with an intraperitoneal  
30 injection of 25 micrograms of each peptide in 0,4 ml of a 1:1 mixture of Freund's Complete Adjuvant (Difco Laboratories, Detroit, MI, USA) and physiological saline (NaCl, 0,15 M). One month later the animals were boosted with an intramuscular injection of 35 micrograms of  
35 antigens in Freund's incomplete adjuvant and saline (1:1) (0,2 ml were distributed into four sites). Three weeks

later the peptides in a dose of 50 micrograms/mouse were administered intravenously and sera were obtained 7 days later.

For the production of EBV transformed B-cells, 5 peripheral blood leukocytes were obtained from healthy control persons. The B-cells were transformed with EBV (Epstein-Barr virus) using standard protocol, and the cell lines were maintained in RPMI 1640, supplemented with 10% FCS (fetal calf serum). An aliquot of cells were stimulated 10 for 12 hours with 10 µg/ml of phytohemagglutinin (PHA) to obtain mitogen-activated T-cells.

Tissue samples were obtained from stillborn fetuses at six months gestational age. Fetal liver, spleen, thymus and lymphnodes were homogenized, the homogenates were 15 cleared with centrifugations (20 000 rpm for 20 minutes) and the samples were used for western blot analysis.

For analysis of polyclonal sera, Elisa and western blot analysis were performed. Microtitre ELISA plates (Maxisorp, Nunc, Roskilde, Denmark) were coated with the 20 peptides (1 micrograms /well in PBS, pH 7,5) at 4°C overnight and blocked with 2 % of BSA in PBS. The plates were then incubated with titrated mouse immune sera and normal (control) sera at room temperature for 4 h. Finally the bound peptide-specific antibodies were detected by use 25 of anti-mouse HRP-labelled immunoglobulins (Dako A/S, Denmark) essentially as previously described [Ovod, V. A. et al., AIDS 6 (1992) 25.34].

For western blotting, tissue homogenates, EBV transformed B-cells or PHA-activated T-cells were boiled 30 for 10 minutes in 2x sample buffer (for tissue homogenates: 100 microliters of homogenate mixed with 100 microliters of sample buffer. For cells: one million cells/100 µl of buffer) and analyzed in western blotting as described in Ovod, V. A. et al., *supra*.

The antisera so produced reacted with the AIR-1- protein low amount in normal fetal spleen, thymus and

lymphonode as well as, in EBV-transformed B-cells and in PHA-activated T-cells. In the ELISA assay towards the immunogenic peptides, all four mice gave a strong reactivity towards the peptide used for the immunization. In the western blotting analysis using either the tissue homogenates or stimulated T-cells or established B-cells, a strong band of approx. 60 kD molecular weight was seen in fetal liver (Fig. 6), while weaker bands of the same size were seen in the other samples.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

- (A) NAME: Kai Krohn et al.
- (B) STREET: Iltarusko, Salmentaantie 751
- (C) CITY: 36450 Salmentaka
- (E) COUNTRY: Finland
- (F) POSTAL CODE (ZIP): none

(ii) TITLE OF INVENTION: Novel Gene

(iii) NUMBER OF SEQUENCES: 26

## (iv) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Floppy disk
- (B) COMPUTER: IBM PC compatible
- (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2036 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION:137..1774
- (D) OTHER INFORMATION:/product= "AIR-1"

## (ix) FEATURE:

- (A) NAME/KEY: mat\_peptide
- (B) LOCATION:137..1771
- (D) OTHER INFORMATION:/product= "AIR-1"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

AGACCGGGGA GACGGGCGGG CGCACAGCCG GCGCGGAGGC CCCACAGCCC CGCCGGGACC	60
CGAGGCCAAG CGAGGGCTG CCAGTGTCCC GGGACCCACC GCGTCCGCC CAGCCCCGGG	120
TCCCCCGCGCC CACCCC ATG GCG ACG GAC GCG GCG CTA CGC CGG CTT CTG Met Ala Thr Asp Ala Ala Leu Arg Arg Leu Leu	169
1                   5                   10	
AGG CTG CAC CGC ACG GAG ATC GCG GTG GCC GTG GAC AGC GGC TTC CCA Arg Leu His Arg Thr Glu Ile Ala Val Ala Val Asp Ser Ala Phe Pro	217
15                 20                 25	
CTG CTG CAC GCG CTG GCT GAC CAC GAC GTG GTC CCC GAG GAC AAG TTT Leu Leu His Ala Leu Ala Asp His Asp Val Val Pro Glu Asp Lys Phe	265
30                 35                 40	

CAG GAG ACG CTT CAT CTG AAG GAA AAG GAG GGC TGC CCC CAG GCC TTC Gln Glu Thr Leu His Leu Lys Glu Lys Glu Gly Cys Pro Gln Ala Phe 45	50	55	313
CAC GCC CTC CTG TCC TGG CTG CTG ACC CAG GAC TCC ACA GCC ATC CTG His Ala Leu Leu Ser Trp Leu Leu Thr Gln Asp Ser Thr Ala Ile Leu 60	65	70	361
GAC TTC TGG AGG GTG CTG TTC AAG GAC TAC AAC CTG GAG CGC TAT GGC Asp Phe Trp Arg Val Leu Phe Lys Asp Tyr Asn Leu Glu Arg Tyr Gly 80	85	90	409
CGG CTG CAG CCC ATC CTG GAC AGC TTC CCC AAA GAT GTG GAC CTC AGC Arg Leu Gln Pro Ile Leu Asp Ser Phe Pro Lys Asp Val Asp Leu Ser 95	100	105	457
CAG CCC CGG AAG GGG AGG AAG CCC CCG GCC GTC CCC AAG GCT TTG GTA Gln Pro Arg Lys Gly Arg Lys Pro Pro Ala Val Pro Lys Ala Leu Val 110	115	120	505
CCG CCA CCC AGA CTC CCC ACC AAG AGG AAG GCC TCA GAA GAG GCT CGA Pro Pro Pro Arg Leu Pro Thr Lys Arg Lys Ala Ser Glu Glu Ala Arg 125	130	135	553
GCT GCC GCG CCA GCA GCC CTG ACT CCA AGG GGC ACC GCC AGC CCA GGC Ala Ala Ala Pro Ala Ala Leu Thr Pro Arg Gly Thr Ala Ser Pro Gly 140	145	150	601
TCT CAA CTG AAG GCC AAG CCC CCC AAG AAG CCG GAG AGC AGC GCA GAG Ser Gln Leu Lys Ala Lys Pro Pro Lys Lys Pro Glu Ser Ser Ala Glu 160	165	170	649
CAG CAG CGC CTT CCA CTC GGG AAC GGG ATT CAG ACC ATG TCA GCT TCA Gln Gln Arg Leu Pro Leu Gly Asn Gly Ile Gln Thr Met Ser Ala Ser 175	180	185	697
GTC CAG AGA GCT GTG GCC ATG TCC TCC GGG GAC GTC CCG GGA GCC CGA Val Gln Arg Ala Val Ala Met Ser Ser Gly Asp Val Pro Gly Ala Arg 190	195	200	745
GGG GCC GTG GAG GGG ATC CTC ATC CAG CAG GTG TTT GAG TCA GGC GGC Gly Ala Val Glu Gly Ile Leu Ile Gln Gln Val Phe Glu Ser Gly Gly 205	210	215	793
TCC AAG AAG TGC ATC CAG GTT GGC GGG GAG TTC TAC ACT CCC AGC AAG Ser Lys Lys Cys Ile Gln Val Gly Glu Phe Tyr Thr Pro Ser Lys 220	225	230	841
TTC GAA GAC TCC GGC AGT GGG AAG AAC AAG GGC CGC AGC AGC AGT GGC Phe Glu Asp Ser Gly Ser Gly Lys Asn Lys Ala Arg Ser Ser Gly 240	245	250	889
CCG AAG CCT CTG GTT CGA GCC AAG GGA GCC CAG GGC GCT GCC CCC GGT Pro Lys Pro Leu Val Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly 255	260	265	937
GGA GGT GAG GCT AGG CTG GGC CAG CAG GGC AGC GTT CCC GCC CCT CTG Gly Gly Glu Ala Arg Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu 270	275	280	985
GCC CTC CCC AGT GAC CCC CAG CTC CAC CAG AAG AAT GAG GAC GAG TGT Ala Leu Pro Ser Asp Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys 285	290	295	1033

GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC TGC TGT GAC GGC TGC CCT Ala Val Cys Arg Asp Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro 300 305 310 315	1081
CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT CCG CTC CGG GAG ATC CCC Arg Ala Phe His Leu Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro 320 325 330	1129
AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG CAG GCA ACA GTC CAG GAG Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu 335 340 345	1177
GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC CAG GAG CCA CCC GTG GAG Val Gln Pro Arg Ala Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu 350 355 360	1225
ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG GGA GAG GAG GTA AGA GGT Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly 365 370 375	1273
CCA CCT GGG GAA CCC CTA GCC GGC ATG GAC ACG ACT CTT GTC TAC AAG Pro Pro Gly Glu Pro Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys 380 385 390 395	1321
CAC CTG CCG GCT CCG CCT TCT GCA GCC CCG CTG CCA GGG CTG GAC TCC His Leu Pro Ala Pro Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser 400 405 410	1369
TCG GCC CTG CAC CCC CTA CTG TGT GTG GGT CCT GAG GGT CAG CAG AAC Ser Ala Leu His Pro Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn 415 420 425	1417
CTG GCT CCT GGT GCG CGT TGC GGG GTG TGC GGA GAT GGT ACG GAC GTG Leu Ala Pro Gly Ala Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val 430 435 440	1465
CTG CGG TGT ACT CAC TGC GCC GCT GCC TTC CAC TGG CGC TGC CAC TTC Leu Arg Cys Thr His Cys Ala Ala Ala Phe His Trp Arg Cys His Phe 445 450 455	1513
CCA GCC GGC ACC TCC CGG CCC GGG ACG GGC CTG CGC TGC AGA TCC TGC Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys 460 465 470 475	1561
TCA GGA GAC GTG ACC CCA GCC CCT GTG GAG GGG GTG CTG GCC CCC AGC Ser Gly Asp Val Thr Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser 480 485 490	1609
CCC GCC CGC CTG GCC CCT GGG CCT GCC AAG GAT GAC ACT GCC AGT CAC Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His 495 500 505	1657
GAG CCC GCT CTG CAC AGG GAT GAC CTG GAG TCC CTT CTG AGC GAG CAC Glu Pro Ala Leu His Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His 510 515 520	1705
ACC TTC GAT GGC ATC CTG CAG TGG GCC ATC CAG AGC ATG GCC CGT CCG Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro 525 530 535	1753
GCG GCC CCC TTC CCC TCC TGA CCCCCAGATGG CCGGGACATG CAGCTCTGAT Ala Ala Pro Phe Pro Ser * 540 545	1804

GAGAGAGTGC TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG AAGCCGCCG GCTGGGATCA 1864  
 AGAAGGGGAC AGCGCCACCT CTTGTCAGTG CTCGGCTGTA AACAGCTCTG TGTTCCTGGG 1924  
 GACACCAGCC ATCATGTGCC TGGAAATTAA ACCCTGCCCC ACTTCTCTAC TCTGGAAGTC 1984  
 CCCGGGAGCC TCTCCTTGCC TGGTGACCTA CTAAAAATAT AAAAATTAGC TG 2036

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 545 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met	Ala	Thr	Asp	Ala	Ala	Leu	Arg	Arg	Leu	Leu	Arg	Leu	His	Arg	Thr
1						5				10				15	
Glu	Ile	Ala	Val	Ala	Val	Asp	Ser	Ala	Phe	Pro	Leu	Leu	His	Ala	Leu
			20							25				30	
Ala	Asp	His	Asp	Val	Val	Pro	Glu	Asp	Lys	Phe	Gln	Glu	Thr	Leu	His
				35			40					45			
Leu	Lys	Glu	Lys	Glu	Gly	Cys	Pro	Gln	Ala	Phe	His	Ala	Leu	Leu	Ser
					55			55				60			
Trp	Leu	Leu	Thr	Gln	Asp	Ser	Thr	Ala	Ile	Leu	Asp	Phe	Trp	Arg	Val
			65				70			75			80		
Leu	Phe	Lys	Asp	Tyr	Asn	Leu	Glu	Arg	Tyr	Gly	Arg	Leu	Gln	Pro	Ile
					85			90				95			
Leu	Asp	Ser	Phe	Pro	Lys	Asp	Val	Asp	Leu	Ser	Gln	Pro	Arg	Lys	Gly
			100					105				110			
Arg	Lys	Pro	Pro	Ala	Val	Pro	Lys	Ala	Leu	Val	Pro	Pro	Arg	Leu	
				115				120				125			
Pro	Thr	Lys	Arg	Lys	Ala	Ser	Glu	Glu	Ala	Arg	Ala	Ala	Ala	Pro	Ala
					130		135				140				
Ala	Leu	Thr	Pro	Arg	Gly	Thr	Ala	Ser	Pro	Gly	Ser	Gln	Leu	Lys	Ala
				145			150			155			160		
Lys	Pro	Pro	Lys	Lys	Pro	Glu	Ser	Ser	Ala	Glu	Gln	Gln	Arg	Leu	Pro
					165			170				175			
Leu	Gly	Asn	Gly	Ile	Gln	Thr	Met	Ser	Ala	Ser	Val	Gln	Arg	Ala	Val
					180			185				190			
Ala	Met	Ser	Ser	Gly	Asp	Val	Pro	Gly	Ala	Arg	Gly	Ala	Val	Glu	Gly
				195			200				205				
Ile	Leu	Ile	Gln	Gln	Val	Phe	Glu	Ser	Gly	Gly	Ser	Lys	Lys	Cys	Ile
					210		215				220				
Gln	Val	Gly	Gly	Glu	Phe	Tyr	Thr	Pro	Ser	Lys	Phe	Glu	Asp	Ser	Gly
				225			230			235			240		

Ser Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val  
 245 250 255  
 Arg Ala Lys Gly Ala Gln Gly Ala Ala Pro Gly Gly Glu Ala Arg  
 260 265 270  
 Leu Gly Gln Gln Gly Ser Val Pro Ala Pro Leu Ala Leu Pro Ser Asp  
 275 280 285  
 Pro Gln Leu His Gln Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp  
 290 295 300  
 Gly Gly Glu Leu Ile Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu  
 305 310 315 320  
 Ala Cys Leu Ser Pro Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg  
 325 330 335  
 Cys Ser Ser Cys Leu Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala  
 340 345 350  
 Glu Glu Pro Arg Pro Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro  
 355 360 365  
 Gly Leu Arg Ser Ala Gly Glu Glu Val Arg Gly Pro Pro Gly Glu Pro  
 370 375 380  
 Leu Ala Gly Met Asp Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro  
 385 390 395 400  
 Pro Ser Ala Ala Pro Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro  
 405 410 415  
 Leu Leu Cys Val Gly Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala  
 420 425 430  
 Arg Cys Gly Val Cys Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His  
 435 440 445  
 Cys Ala Ala Ala Phe His Trp Arg Cys His Phe Pro Ala Gly Thr Ser  
 450 455 460  
 Arg Pro Gly Thr Gly Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr  
 465 470 475 480  
 Pro Ala Pro Val Glu Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala  
 485 490 495  
 Pro Gly Pro Ala Lys Asp Asp Thr Ala Ser His Glu Pro Ala Leu His  
 500 505 510  
 Arg Asp Asp Leu Glu Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile  
 515 520 525  
 Leu Gln Trp Ala Ile Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro  
 530 535 540  
 Ser \*  
 545

## (2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1545 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 237..1283
- (C) OTHER INFORMATION:/product= "AIR-2"

(ix) FEATURE:

- (A) NAME/KEY: mat\_peptide
- (B) LOCATION: 237..1280
- (C) OTHER INFORMATION:/product= "AIR-2"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AGAGAAAGTG AGGTCTTCTC AGGCTCTTAA GAGCATGGCG TTTGGTCCAG GCTGTACCCG	60
CTGCTCTCAG CTGGGCCCGT GGGTGGGCCG GGCGCCCCCTG CTATAGCCAG GAGGTCAAGG	120
ATCCACTGGG AATGCCATGC TCATCTTCG TCCCCAGCAT GGTTTCTTAA TGGGGTAGAA	180
GCAGGGTCGGG AGAGACCTCC CTGGGCCTGG CCCCACTGCC CTGTGAGGAA GGGTTC	236
ATG TGG TTG GTG TAC AGT TCC GGG GCC CCT GGA ACG CAG CAG CCT GCA Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala	284
1 5 10 15	
AGA AAC CGG GTT TTC CCA ATA GGG ATG GCC CCG GGG GGT GTC TGT Arg Asn Arg Val Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys	332
20 25 30	
TGG AGA CCA GAT GGA TGG GGA ACA GGT GGT CAG GGC AGA ATT TCA GGC Trp Arg Pro Asp Gly Trp Gly Thr Gly Gln Gly Arg Ile Ser Gly	380
35 40 45	
CCT GGC AGC ATG GGA GCA GGG CAG AGA CTG GGG AGT TCA GGT ACC CAG Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln	428
50 55 60	
AGA TGC TGC TGG GGG AGC TGT TTT GGG AAG GAG GTG GCT CTC AGG AGG Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg	476
65 70 75 80	
GTG CTG CAC CCC AGC CCA GTC TGC ATG GGC GTC TCT TGC CTG TGC CAG Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln	524
85 90 95	
AAG AAT GAG GAC GAG TGT GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile	572
100 105 110	
TGC TGT GAC GGC TGC CCT CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro	620
115 120 125	

CCG CTC CGG GAG ATC CCC AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu 130 135 140	668
CAG GCA ACA GTC CAG GAG GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro 145 150 155 160	716
CAG GAG CCA CCC GTG GAG ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala 165 170 175	764
GGA GAG GAG GTA AGA GGT CCA CCT GGG GAA CCC CTA GCC GGC ATG GAC Gly Glu Val Arg Gly Pro Pro Gly Glu Pro Leu Ala Gly Met Asp 180 185 190	812
ACG ACT CTT GTC TAC AAG CAC CTG CCG GCT CCG CCT TCT GCA GCC CCG Thr Thr Leu Val Tyr Lys His Leu Pro Ala Pro Pro Ser Ala Ala Pro 195 200 205	860
CTG CCA GGG CTG GAC TCC TCG GCC CTG CAC CCC CTA CTG TGT GTG GGT Leu Pro Gly Leu Asp Ser Ser Ala Leu His Pro Leu Leu Cys Val Gly 210 215 220	908
CCT GAG GGT CAG CAG AAC CTG GCT CCT GGT GCG CGT TGC GGG GTG TGC Pro Glu Gly Gln Gln Asn Leu Ala Pro Gly Ala Arg Cys Gly Val Cys 225 230 235 240	956
GGA GAT GGT ACG GAC GTG CTG CGG TGT ACT CAC TGC GCC GCT GCC TTC Gly Asp Gly Thr Asp Val Leu Arg Cys Thr His Cys Ala Ala Ala Phe 245 250 255	1004
CAC TGG CGC TGC CAC TTC CCA GCC GGC ACC TCC CGG CCC GGG ACG GGC His Trp Arg Cys His Phe Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly 260 265 270	1052
CTG CGC AGA TCC TGC TCA GGA GAC GTG ACC CCA GCC CCT GTG GAG Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr Pro Ala Pro Val Glu 275 280 285	1100
GGG GTG CTG GCC CCC AGC CCC GCC CGC CTG GCC CCT GGG CCT GCC AAG Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys 290 295 300	1148
GAT GAC ACT GCC AGT CAC GAG CCC GCT CTG CAC AGG GAT GAC CTG GAG Asp Asp Thr Ala Ser His Glu Pro Ala Leu His Arg Asp Asp Leu Glu 305 310 315 320	1196
TCC CTT CTG AGC GAG CAC ACC TTC GAT GGC ATC CTG CAG TGG GCC ATC Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile 325 330 335	1244
CAG AGC ATG GCC CGT CCG GCG CCC TTC CCC TCC TGA CCCCCAGATGG Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro Ser *	1293
340 345	
CGGGGACATG CAGCTCTGAT GAGAGAGTGC TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG	1353
AAGCCGGCCG GCTGGGATCA AGAAGGGAC AGCGCCACCT CTTGTCAGTG CTCGGCTGTA	1413
AACAGCTCTG TGTTTCTGGG GACACCAGCC ATCATGTGCC TGGAAATTAA ACCCTGCC	1473

ACTTCTCTAC TCTGGAAAGTC CCCGGGAGCC TCTCCTTGCC TGGTGACCTA CTAAAAATAT 1533  
 AAAAATTAGC TG 1545

## (2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 348 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met	Trp	Leu	Val	Tyr	Ser	Ser	Gly	Ala	Pro	Gly	Thr	Gln	Gln	Pro	Ala
1					5				10					15	
Arg	Asn	Arg	Val	Phe	Phe	Pro	Ile	Gly	Met	Ala	Pro	Gly	Gly	Val	Cys
					20				25				30		
Trp	Arg	Pro	Asp	Gly	Trp	Gly	Thr	Gly	Gly	Gln	Gly	Arg	Ile	Ser	Gly
					35			40				45			
Pro	Gly	Ser	Met	Gly	Ala	Gly	Gln	Arg	Leu	Gly	Ser	Ser	Gly	Thr	Gln
					50		55		60						
Arg	Cys	Cys	Trp	Gly	Ser	Cys	Phe	Gly	Lys	Glu	Val	Ala	Leu	Arg	Arg
					65		70		75				80		
Val	Leu	His	Pro	Ser	Pro	Val	Cys	Met	Gly	Val	Ser	Cys	Leu	Cys	Gln
					85			90				95			
Lys	Asn	Glu	Asp	Glu	Cys	Ala	Val	Cys	Arg	Asp	Gly	Gly	Glu	Leu	Ile
					100			105				110			
Cys	Cys	Asp	Gly	Cys	Pro	Arg	Ala	Phe	His	Leu	Ala	Cys	Leu	Ser	Pro
					115			120				125			
Pro	Leu	Arg	Glu	Ile	Pro	Ser	Gly	Thr	Trp	Arg	Cys	Ser	Ser	Cys	Leu
					130		135			140					
Gln	Ala	Thr	Val	Gln	Glu	Val	Gln	Pro	Arg	Ala	Glu	Glu	Pro	Arg	Pro
					145		150			155			160		
Gln	Glu	Pro	Pro	Val	Glu	Thr	Pro	Leu	Pro	Pro	Gly	Leu	Arg	Ser	Ala
					165			170				175			
Gly	Glu	Glu	Val	Arg	Gly	Pro	Pro	Gly	Glu	Pro	Leu	Ala	Gly	Met	Asp
					180			185				190			
Thr	Thr	Leu	Val	Tyr	Lys	His	Leu	Pro	Ala	Pro	Pro	Ser	Ala	Ala	Pro
					195			200				205			
Leu	Pro	Gly	Leu	Asp	Ser	Ser	Ala	Leu	His	Pro	Leu	Leu	Cys	Val	Gly
					210			215				220			
Pro	Glu	Gly	Gln	Gln	Asn	Leu	Ala	Pro	Gly	Ala	Arg	Cys	Gly	Val	Cys
					225			230			235		240		
Gly	Asp	Gly	Thr	Asp	Val	Leu	Arg	Cys	Thr	His	Cys	Ala	Ala	Ala	Phe
					245			250				255			

His Trp Arg Cys His Phe Pro Ala Gly Thr Ser Arg Pro Gly Thr Gly  
 260 265 270  
 Leu Arg Cys Arg Ser Cys Ser Gly Asp Val Thr Pro Ala Pro Val Glu  
 275 280 285  
 Gly Val Leu Ala Pro Ser Pro Ala Arg Leu Ala Pro Gly Pro Ala Lys  
 290 295 300  
 Asp Asp Thr Ala Ser His Glu Pro Ala Leu His Arg Asp Asp Leu Glu  
 305 310 320  
 Ser Leu Leu Ser Glu His Thr Phe Asp Gly Ile Leu Gln Trp Ala Ile  
 325 330 335  
 Gln Ser Met Ala Arg Pro Ala Ala Pro Phe Pro Ser \*  
 340 345

## (2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1463 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ix) FEATURE:  
 (A) NAME/KEY: CDS  
 (B) LOCATION:237..1001  
 (D) OTHER INFORMATION:/product= "AIR-3"  
 (ix) FEATURE:  
 (A) NAME/KEY: mat\_peptide  
 (B) LOCATION:237..998  
 (D) OTHER INFORMATION:/product= "AIR-3"

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AGAGAAAGTG AGGTCTTCTC AGGCTTTAA GAGCATGGCG TTTGGTCCAG GCTGTACCCG	60
CTGCTCTCAG CTGGGCCCGT GGGTGGGCCG GGCGCCCCCTG CTATAGCCAG GAGGTCAAGG	120
ATCCACTGGG AATGCCATGC TCATCTTCG TCCCCAGCAT GGTTTCTTAA TGGGGTAGAA	180
GCAGGTCGGG AGAGACCTCC CTGGGCCTGG CCCCAGTGC CTGTGAGGAA GGGTTC	236
ATG TGG TTG GTG TAC AGT TCC GGG GCC CCT GGA ACG CAG CAG CCT GCA Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala 1 5 10 15	284
AGA AAC CGG GTT TTC TTC CCA ATA GGG ATG GCC CCG GGG GGT GTC TGT Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Val Cys 20 25 30	332
TGG AGA CCA GAT GGA TGG GGA ACA GGT GGT CAG GGC AGA ATT TCA GGC Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly 35 40 45	380
CCT GGC AGC ATG GGA GCA GGG CAG AGA CTG GGG AGT TCA GGT ACC CAG Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln 50 55 60	428

AGA TGC TGC TGG GGG AGC TGT TTT GGG AAG GAG GTG GCT CTC AGG AGG Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg 65 70 75 80	476
GTG CTG CAC CCC AGC CCA GTC TGC ATG GGC GTC TCT TGC CTG TGC CAG Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln 85 90 95	524
AAG AAT GAG GAC GAG TGT GCC GTG TGT CGG GAC GGC GGG GAG CTC ATC Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile 100 105 110	572
TGC TGT GAC GGC TGC CCT CGG GCC TTC CAC CTG GCC TGC CTG TCC CCT Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro 115 120 125	620
CCG CTC CGG GAG ATC CCC AGT GGG ACC TGG AGG TGC TCC AGC TGC CTG Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu 130 135 140	668
CAG GCA ACA GTC CAG GAG GTG CAG CCC CGG GCA GAG GAG CCC CGG CCC Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro 145 150 155 160	716
CAG GAG CCA CCC GTG GAG ACC CCG CTC CCC CCG GGG CTT AGG TCG GCG Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala 165 170 175	764
GGA GAG GAG CCC CGC TGC CAG GGC TGG ACT CCT CGG CCC TGC ACC CCC Gly Glu Pro Arg Cys Gln Gly Trp Thr Pro Arg Pro Cys Thr Pro 180 185 190	812
TAC TGT GTG TGG GTC CTG AGG GTC AGC AGA ACC TGG CTC CTG GTG CGC Tyr Cys Val Trp Val Leu Arg Val Ser Arg Thr Trp Leu Leu Val Arg 195 200 205	860
GTT GCG GGG TGT GCG GAG ATG GTA CGG ACG TGC TGC GGT GTA CTC ACT Val Ala Gly Cys Ala Glu Met Val Arg Thr Cys Cys Gly Val Leu Thr 210 215 220	908
GCG CCG CTG CCT TCC ACT GGC GCT GCC ACT TCC CAG CCG GCA CCT CCC Ala Pro Leu Pro Ser Thr Gly Ala Ala Thr Ser Gln Pro Ala Pro Pro 225 230 235 240	956
GGC CCG GGA CGG GCC TGC GCT GCA GAT CCT GCT CAG GAG ACG TGA Gly Pro Gly Arg Ala Cys Ala Ala Asp Pro Ala Gln Glu Thr * 245 250 255	1001
CCCCAGCCCC TGTGGAGGGG GTGCTGGCCC CCAGCCCCGC CCGCCTGGCC CCTGGGCCCTG	1061
CCAAGGATGA CACTGCCAGT CACGAGCCCG CTCTGCACAG GGATGACCTG GAGTCCCTTC	1121
TGAGCGAGCA CACCTTCGAT GGCATCCTGC AGTGGGCCAT CCAGAGCATG GCCCGTCCGG	1181
CGGCCCCCTT CCCCTCCTGA CCCCAGATGG CCGGGACATG CAGCTCTGAT GAGAGAGTGC	1241
TGAGAAGGAC ACCTCCTTCC TCAGTCCTGG AAGCCGGCCG GCTGGGATCA AGAAGGGGAC	1301
AGCGCCACCT CTTGTCAGTG CTCGGCTGTA AACAGCTCTG TGTTTCTGGG GACACCAGCC	1361
ATCATGTGCC TGGAAATTAA ACCCTGCCCTT ACCTCTCTAC TCTGGAAGTC CCCGGGAGCC	1421
TCTCCTTGCC TGGTGACCTA CTAAAAATAT AAAAATTAGC TG	1463

## (2) INFORMATION FOR SEQ ID NO: 6:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 254 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: protein

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Met Trp Leu Val Tyr Ser Ser Gly Ala Pro Gly Thr Gln Gln Pro Ala  
 1 5 10 15

Arg Asn Arg Val Phe Phe Pro Ile Gly Met Ala Pro Gly Gly Val Cys  
 20 25 30

Trp Arg Pro Asp Gly Trp Gly Thr Gly Gly Gln Gly Arg Ile Ser Gly  
 35 40 45

Pro Gly Ser Met Gly Ala Gly Gln Arg Leu Gly Ser Ser Gly Thr Gln  
 50 55 60

Arg Cys Cys Trp Gly Ser Cys Phe Gly Lys Glu Val Ala Leu Arg Arg  
 65 70 75 80

Val Leu His Pro Ser Pro Val Cys Met Gly Val Ser Cys Leu Cys Gln  
 85 90 95

Lys Asn Glu Asp Glu Cys Ala Val Cys Arg Asp Gly Gly Glu Leu Ile  
 100 105 110

Cys Cys Asp Gly Cys Pro Arg Ala Phe His Leu Ala Cys Leu Ser Pro  
 115 120 125

Pro Leu Arg Glu Ile Pro Ser Gly Thr Trp Arg Cys Ser Ser Cys Leu  
 130 135 140

Gln Ala Thr Val Gln Glu Val Gln Pro Arg Ala Glu Glu Pro Arg Pro  
 145 150 155 160

Gln Glu Pro Pro Val Glu Thr Pro Leu Pro Pro Gly Leu Arg Ser Ala  
 165 170 175

Gly Glu Glu Pro Arg Cys Gln Gly Trp Thr Pro Arg Pro Cys Thr Pro  
 180 185 190

Tyr Cys Val Trp Val Leu Arg Val Ser Arg Thr Trp Leu Leu Val Arg  
 195 200 205

Val Ala Gly Cys Ala Glu Met Val Arg Thr Cys Cys Gly Val Leu Thr  
 210 215 220

Ala Pro Leu Pro Ser Thr Gly Ala Ala Thr Ser Gln Pro Ala Pro Pro  
 225 230 235 240

Gly Pro Gly Arg Ala Cys Ala Ala Asp Pro Ala Gln Glu Thr \*  
 245 250 255

## (2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GATGACACTG CCAGTCACGA

20

## (2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTTCCCGAGT GGAAGGCGCT GC

22

## (2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

AGGGGACAGG CAGGCCAGGT

20

## (2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

GAGTTCAGGT ACCCAGAGAT GCTG

24

## (2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

CTCGCTCAGA AGGGACTCCA

20

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGATTCAGAC CATGTCAGCT TCA

23

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

GTGCTGTTCA AGGACTACAA C

21

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TGGATGAGGA TCCCCTCCAC G

21

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

TGCAGGCTGT GGGAACTCCA

20

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

AGAAAAAAGAG CTGTACCCCTG TG

22

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

TGCAAGGAAG AGGGGCGTCA GC

22

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

TCCACCACAA GCCGAGGAGA T

21

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

ACGGGCTCCT CAAACACCAC T

21

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

TGGAGATGGG CAGGCCGCAG GGTG

24

## (2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
24

CAGTCCAGCT GGGCTGAGCA GGTG

## (2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:  
24

GCGGCTCCAA GAAGTGCATC CAGG

## (2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
24

CTCCACCCCTG CAAGGAAGAG GGGC

## (2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

Thr	Leu	His	Leu	Lys	Glu	Lys	Gly	Cys	Pro	Gln	Ala	Phe	His
													15
1				5									

## (2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Gly Lys Asn Lys Ala Arg Ser Ser Ser Gly Pro Lys Pro Leu Val  
1 5 10 15

**Claims**

1. An isolated DNA sequence characterized by comprising the sequence id. no. 1 or a fragment or variant thereof, or an isolated DNA sequence hybridizable thereto, the 5 DNA sequence being associated with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED).
2. An isolated DNA sequence according to claim 1, characterized in that it includes a gene defect responsible 10 for APECED.
3. A DNA sequence according to claim 1, characterized by having the sequence according to sequence id. no 1 or a fragment thereof having the sequence according to sequence id. no 3 or sequence id. no 5.
4. A protein characterized by comprising the amino acid sequence id. no. 2 or a fragment or variant thereof, the protein being associated with autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy 15 (APECED).
5. A protein according to claim 4 characterized by 20 having the amino acid sequence id. no. 2, or a fragment thereof having the sequence according to sequence id. no. 4, or a fragment thereof having the sequence according to sequence id. no 6.
6. A protein according to claim 4 or 5 characterized by 25 having distinct structural motifs, including the PHD finger motif (PHD), the LXXLL motif (L), proline-rich region (PRR), and cystein-rich region (CRR).
7. A method for the diagnosis of autoimmune poly- 30 endocrinopathy-candidiasis-ectodermal dystrophy (APECED) characterized by detecting in a biological specimen the presence of a DNA sequence comprising the sequence id. no. 1 or a functional fragment or variant thereof, or an isolated DNA-sequence hybridizable thereto, the DNA 35 sequence being associated with APECED.

8. A method according to claim 7, characterized in that the DNA sequence includes a gene defect responsible for APECED.

9. A method according to claim 8, characterized in 5 that the gene defect to be detected includes a "C" to "T" transition resulting in the "Arg" to "Stop" nonsense mutation at amino acid position 257 and/or a "A" to "G" transversion resulting in the "Lys" to "Glu" missense mutation at amino acid position 42.

10. A method according to any one of claims 7 to 9, characterized in that DNA techniques are used for the detection.

11. A method according to any one of claims 7 to 10, characterized in that the detection takes advantage of 15 TaqI or another enzyme cleaving at recognition site 5'-TCGA-3' digestion.

12. A method for the diagnosis of autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) characterized by detecting in a biological 20 specimen the presence or the absence of a protein comprising the sequence id. no. 1, or a fragment thereof having the sequence according to sequence id. no. 4, or a fragment therof having the sequence according to sequence id. no 6, the protein being associated with APECED.

25 13. The use of the DNA sequence according to any one of claims 1 to 3 in the diagnosis of APECED.

14. The use of the protein according to any one of claims 4 to 6 in the diagnosis of APECED.

15. The use of the DNA sequence according to any one 30 of claims 1 to 3 for the preparation of a medicament useful in a gene therapy method of APECED.

16. The use of the DNA sequence according to any one of claims 1 to 3 in the treatment of APECED.

## (57) Abstract

The present invention relates to a novel gene, a novel protein encoded by said gene, a mutated form of the gene and to diagnostic and therapeutic uses of the gene or a 5 mutated form thereof. More specifically, the present invention relates to a novel gene defective in autoimmune polyendocrinopathy syndrome type I (APS I), also called autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED) (MIM No. 240,300).

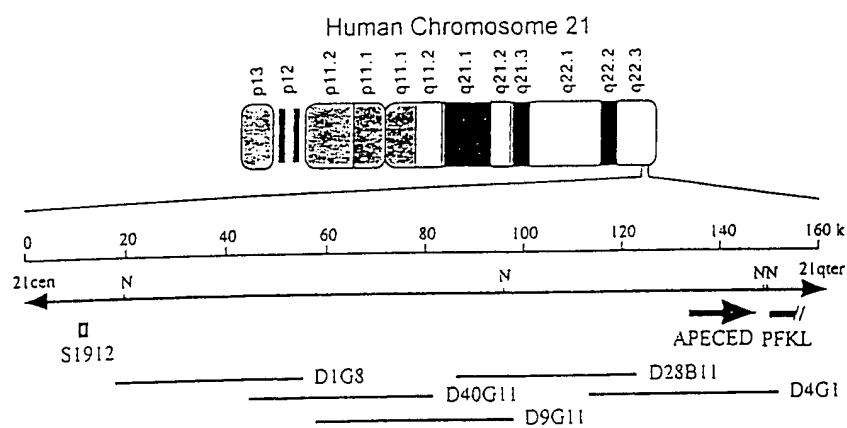


Fig. 1

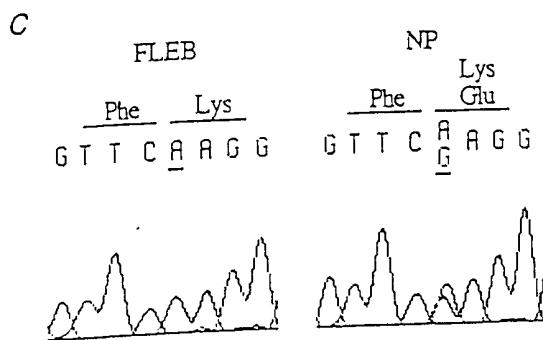
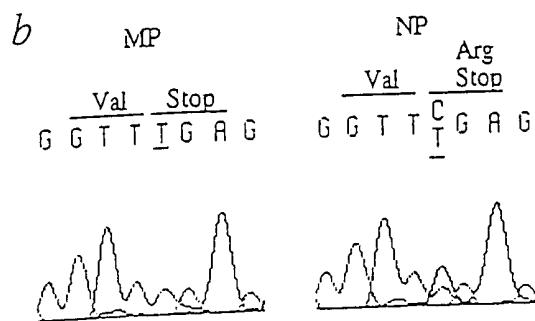
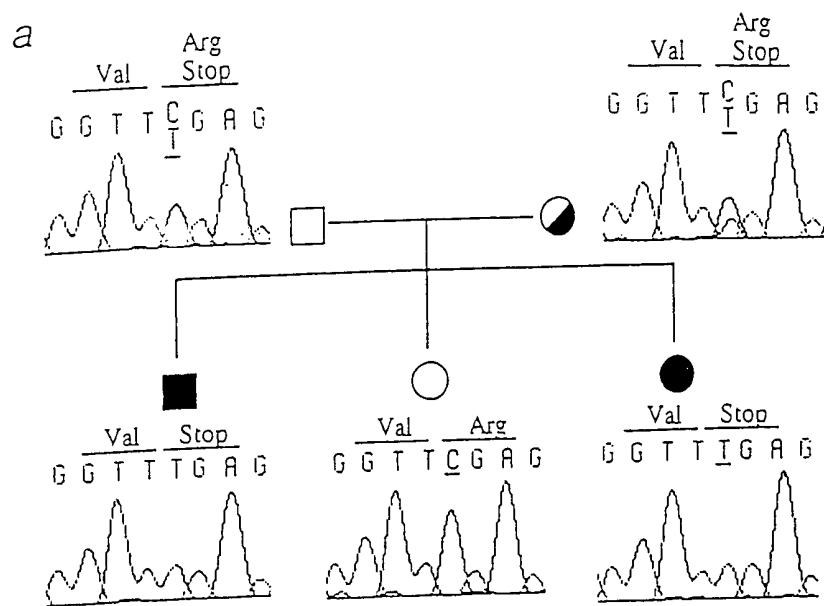


Fig. 3

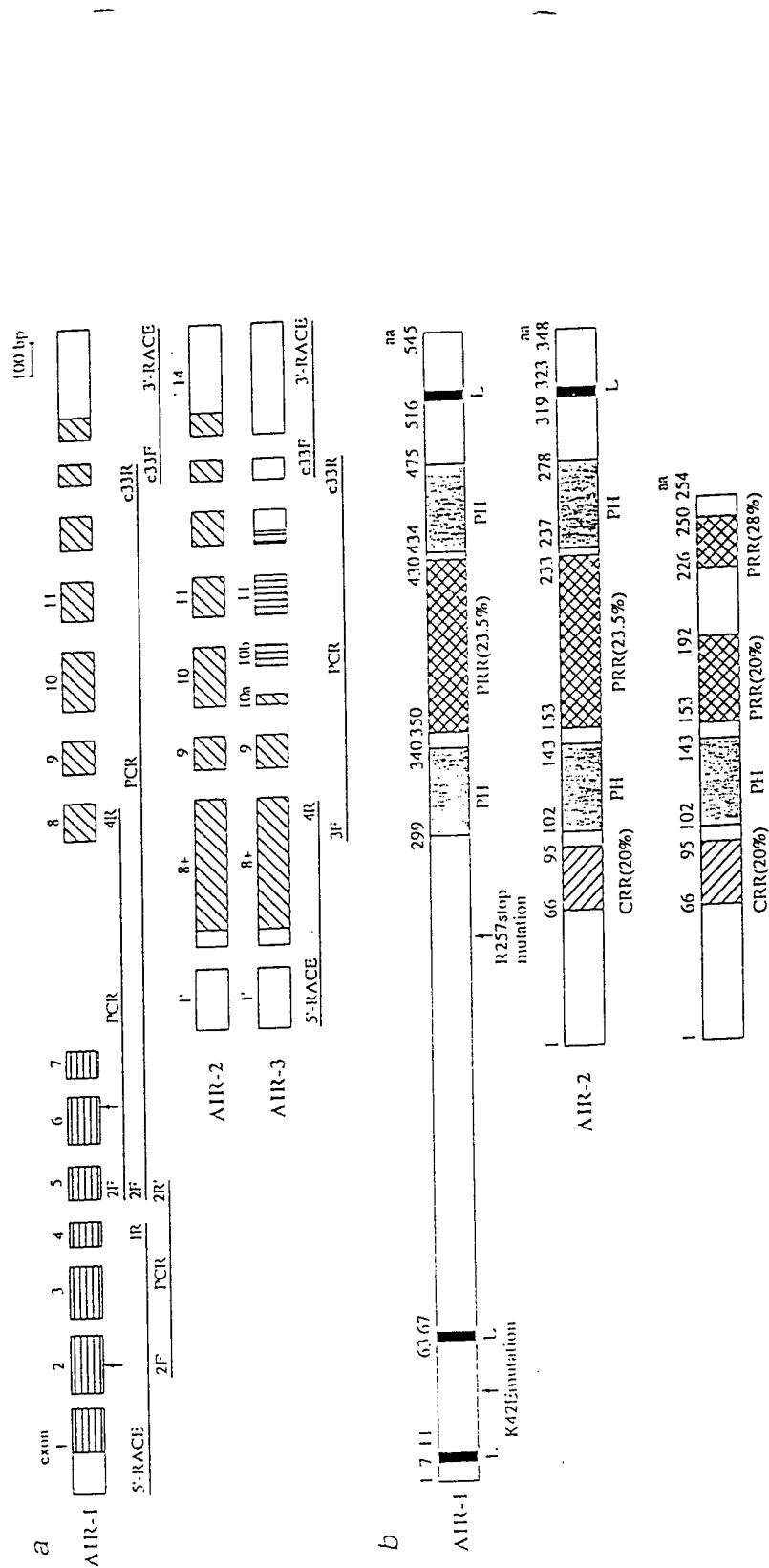


Fig. 2

1 2 3 4 5 6



Fig. 6

AIR-1: 299 CAVCRDGELICCDGCPRAFHACLSPLREIPSGTWRCSSC 340  
AIR-1: 434 .G..G..TDVLR.TH.AA...WR.HF.AGTSR.GTGL..R.. 475  
Mi-2 : 373 .E..QQ...I.L..T....Y.MV..D.DMEKA.E.K.S.PH. 414  
Mi-2 : 452 .R..K.....T..SSY.IH..N...P...N.E.L.PR. 493  
TIF1 : 791 ....QN.....EK..KV...S.HV.T.TNF...E.I.TF. 832  
consensus C C C C H C C C

Fig. 5

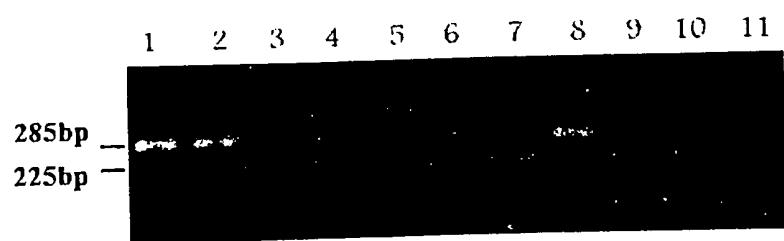


Fig.4